Systems and Computers in Japan, Vol. 37, No. 4, 2006
Translated from Denshi Joho Tsushin Gakkai Ronbunshi, Vol. J84-D-II, No. 10, October 2001, pp. 2201–2211

Background Subtraction under Varying Illumination

Takashi Matsuyama, Toshikazu Wada, Hitoshi Habe,* and Kazuya Tanahashi†

Graduate School of Informatics, Kyoto University, Kyoto, 606-8501 Japan

SUMMARY

Background subtraction is widely used as an effective method for detecting moving objects in a video image. However, background subtraction requires a prerequisite in that image variation cannot be observed, and the range of application is limited. Proposed in this research paper is a method for detecting moving objects by using background subtraction that can be applied to cases in which the image has varied due to varying illumination. This method is based on two object detection methods that are based on different lines of thinking. One method compares the background image and the observed image using invariant features of illumination. The other method estimates the illumination conditions of the observed image and normalizes the brightness before carrying out background subtraction. These two methods are complementary, and highly precise detection results can be obtained by ultimately integrating the detection results of both methods. © 2006 Wiley Periodicals, Inc. Syst Comp Jpn, 37(4): 77–88, 2006; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/scj.10166

Key words: background subtraction; object detection; varying illumination; normalized vector distance; eigenimage analysis.

*Currently with Mitsubishi Electric Corporation.
†Currently with NTT Data Corporation.

1. Introduction

Background subtraction is used as an effective method for detecting moving objects in a video image. However, background subtraction requires a prerequisite in that image variation cannot be observed, and the range of application is limited.

Background scene variation in an image is commonly caused by (1) camera motion, (2) varying illumination, and (3) movement of objects in the background (swaying of trees, CRT flickering, and the like).

The authors have proposed a fixed-viewpoint pan-tilt-zoom camera [1, 2] designed to handle camera movement. Since the viewpoint does not move due to rotation and zooming when this camera is used, the viewpoint and zoom control is performed as background subtraction is being carried out.

Many techniques have been proposed in relation to varying illumination and object movement in the background. References 3 and 4 model the variation in the pixel values of a background image by using a probability distribution to detect the pixels corresponding to the moving object. References 6 and 7 adaptively regenerate a background image with respect to varying illumination and the varying appearance of objects in the background. Toyama and colleagues have recently proposed a multilevel background subtraction method [8]. In this technique, the constraint conditions in the spatial direction and the temporal direction are integrated, and the accuracy of background subtraction is improved for each pixel via a Wiener filter.

Proposed in this research paper is a robust background subtraction method under varying illumination. A precondition is that all of the objects in a background scene

are stationary, but it is possible to consider that a technique for handling a wide range of background variation can be implemented by combining background subtraction methods based on the use of spatial and temporal continuity, as shown in Ref. 8.

In this research paper, two detection methods based on different lines of thinking will first be introduced. One method compares the background image and the observed image by using invariant features of illumination. The other method estimates the illumination conditions of the observed image and normalizes the luminance before carrying out background subtraction. These two methods are complementary, and the authors propose a method for obtaining highly precise detection results by ultimately integrating the detection results of both methods. Finally, the effectiveness of this method under varying illumination is shown by way of a performance evaluation test.

2. Background Subtraction Using Invariable Properties in Illumination

2.1. Normalized vector distance

Normalized vector distance [5] (NVD) is a feature that is not easily affected by varying illumination. In order to calculate normalized vector distance, an image is first divided into blocks of $N \times N$ pixels, and each block is expressed in terms of an N^2-dimensional vector. As used herein, the elements of the vector correspond to the luminance value in each of the blocks. The vectors $\boldsymbol{i}_{(u,v)}$ and $\boldsymbol{b}_{(u,v)}$ correspond to blocks in the same position in the observed image and the background image, respectively. These vectors will hereafter be referred to as "image vectors." It follows then that the normalized vector distance is given by the following expression (Fig. 1):

$$ND(\boldsymbol{i}_{(u,v)}) = \left| \frac{\boldsymbol{i}_{(u,v)}}{|\boldsymbol{i}_{(u,v)}|} - \frac{\boldsymbol{b}_{(u,v)}}{|\boldsymbol{b}_{(u,v)}|} \right| \tag{1}$$

In the expression, the terms in | | represent the magnitude of the vectors.

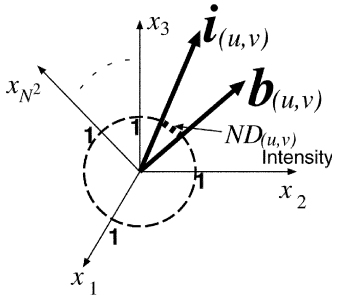


Fig. 1. Normalized vector distance.

From the definition it is clear that $ND(\boldsymbol{i}_{(u,v)})$ is not affected by the magnitude of the input image vector, that is, the uniform variation of the luminance in a block. In real terms, normalized vector distance and normalized cross-correlation satisfy the following relational expression:

$$ND(\boldsymbol{i}_{(u,v)}) = \sqrt{2(1-\cos\theta)} \tag{2}$$

In the expression, θ is the angle between the vectors $\boldsymbol{i}_{(u,v)}$ and $\boldsymbol{b}_{(u,v)}$. The normalized cross-correlation calculated between the blocks is none other than cos θ.

However, $\boldsymbol{i}_{(u,v)}$ commonly contains noise, and since ratios are calculated for the normalized vector distance, the effect of noise increases when the magnitude of the input image vector is small, and the value of the normalized vector distance becomes unstable.

The method in this research paper handles the problem in the following manner by

(1) adaptively varying the threshold value when detecting moving objects under varying illumination on the basis of an analysis of the statistical characteristics of the normalized vector distance, and

(2) enhancing the normalized vector distance by evaluating the spatial properties of variation within a block.

2.2. Determining an adaptive threshold value

In the observed image, $\boldsymbol{i}^B_{(u,v)}$ is the image vector corresponding to the background scene, and the effect of noise $\boldsymbol{n}$ is postulated to be additive, as shown below:

$$\boldsymbol{i}^B_{(u,v)} = \widetilde{\boldsymbol{i}^B_{(u,v)}} + \boldsymbol{n} \tag{3}$$

In the expression, $\widetilde{\boldsymbol{i}^B_{(u,v)}} = \alpha \boldsymbol{b}_{(u,v)}$ is a parameter that expresses uniform luminance variation within a block due to illumination variation. The elements of $\boldsymbol{n}$ follow an independent normal distribution in which the mean is 0 and the standard deviation is σ.

Therefore, the theorem shown below can be derived with regard to the normalized vector distance (refer to the Appendix).

[Theorem 1] The mean m_{ND}, and variance v_{ND} of $ND(\boldsymbol{i}^B_{(u,v)})$ can be approximated as follows using $|\widetilde{\boldsymbol{i}^B_{(u,v)}}|$, σ, N:

$$m_{ND}(\boldsymbol{i}^B_{(u,v)}) = \frac{\Gamma\left(\frac{N^2}{2}\right)}{\Gamma\left(\frac{N^2-1}{2}\right)} \frac{\sqrt{2}\sigma}{|\widetilde{\boldsymbol{i}^B_{(u,v)}}|}$$

$$v_{ND}(\boldsymbol{i}^B_{(u,v)}) = \left[\frac{\Gamma\left(\frac{N^2+1}{2}\right)}{\Gamma\left(\frac{N^2-1}{2}\right)} - \frac{\Gamma\left(\frac{N^2}{2}\right)^2}{\Gamma\left(\frac{N^2-1}{2}\right)^2} \right] \frac{2\sigma^2}{|\widetilde{\boldsymbol{i}^B_{(u,v)}}|^2} \tag{4}$$

In the formula, Γ( ) is a gamma function.

The theorem shows that "the effect of noise on the normalized vector distance is determined solely by luminance $\tilde{\boldsymbol{i}}^{B}_{(u,v)}$ " in conditions in which ideal illumination variation and noise are observed. Therefore, as described below, if the luminance $\tilde{\boldsymbol{i}}^{B}_{(u,v)}$ of the background can be estimated, highly precise detection can be brought about with consideration for the effect of noise. The method of estimating the luminance is described in the next section. In the following discussion, the $m_{ND}\,\boldsymbol{i}^{B}_{(u,v)}$ and $v_{ND}\,\boldsymbol{i}^{B}_{(u,v)}$ in each block of the observed image are assumed to be known by estimation.

Figure 2 shows the method of determining the threshold value in which this theorem is used.

In the diagram, the horizontal axis is $\tilde{\boldsymbol{i}}^{B}_{(u,v)} = \alpha \boldsymbol{b}_{(u,v)}$, and the vertical axis is $ND(\boldsymbol{i}_{(u,v)})$. The solid line in the lower portion of the diagram is the mean m_{ND} derived from Eq. (4), and the points in the vicinity of the line are the mean values given by the actual image. In the calculation, the magnitude of the blocks is empirically determined, and $N = 16$. The results confirm that the theorem is valid. The dot-dash line in the center is $m_{ND} + \sqrt{v_{ND}}$ and the thickly dotted line is $m_{ND} + 2\sqrt{v_{ND}}$. The diagram shows the situation in which a moving object is detected with $m_{ND} + 2\sqrt{v_{ND}}$ as the threshold value. In this manner, adaptive background subtraction can be brought about* by varying the threshold value in each of the blocks in accordance with the luminance of the block $|\tilde{\boldsymbol{i}}^{B}_{(u,v)}|$.

2.3. Integration with spatial characteristics

Misdetections caused by noise can be reduced by adaptively varying the threshold value, as shown in Fig. 2. However, this process simply reduces sensitivity in dark areas. This is a common problem in image processing in which color differences and other calculations of ratios are used in addition to block correlation, and in order to solve this problem, information other than brightness must be included.

In this research, detection accuracy is ensured through the normalized vector distance by giving consideration to the spatial characteristics of the varying brightness value within a block. More specifically, variation due to noise and variation due to a moving object are identified based on the spatial characteristics of the variation in the brightness value within a block.

The following postulates are introduced to characterize the spatial configuration of the variation in the brightness value within a block.

[Postulate 1] Brightness variation due to noise is independently and uniformly distributed within a block.

*The standard deviation σ of the noise component is calculated in advance for each imaging system.

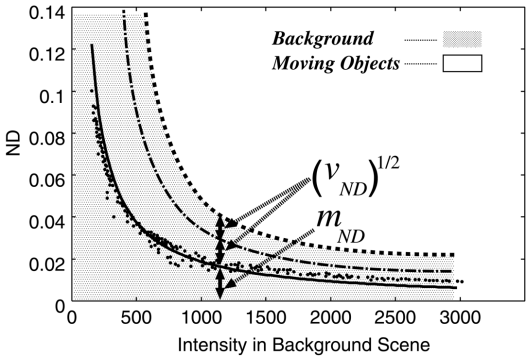


Fig. 2. Adaptive threshold determination based on the statistical properties of NVD.

The brightness variation due to a moving object is concentrated and distributed in a specific area within a block.

The measure for evaluation such as the one below is defined based on this postulate. First, the blocks $B_{(u,v)}$ and $I_{(u,v)}$ in the same position of the background image, and the observed image are each divided into small windows, as shown in Fig. 3. Let m be the number of small windows within a single block (m = 5 in Fig. 3). The variables $\boldsymbol{w}^{j}_{B_{(u,v)}}$ and $\boldsymbol{w}^{j}_{I_{(u,v)}}$ represent the j-th window inside of $B_{(u,v)}$ and $I_{(u,v)}$, respectively.

The dispersion of the normalized vector distance calculated for the small windows inside the block is expressed as

$$VND(\boldsymbol{i}_{(u,v)}) = \frac{1}{m}\sum_{j=1}^{m}\left(C^{j}_{(u,v)} - \overline{C_{(u,v)}}\right)^{2} \quad (5)$$

In the formula, $C^{j}_{(u,v)}$ is the normalized vector distance between the small windows $\boldsymbol{w}^{j}_{B_{(u,v)}}$ and $\boldsymbol{w}^{j}_{I(u,v)}$, and $\overline{C_{(u,v)}} = 1/m\,\Sigma_{j=1}^{m} C^{j}_{(u,v)}$ is the mean value within the block. The spatial characteristics of the variation within the block can be analyzed by using $VND(i_{(u,v)})$. The spatial characteristics of the variation appearing in the block can be classified into three types, as shown in Fig. 4.

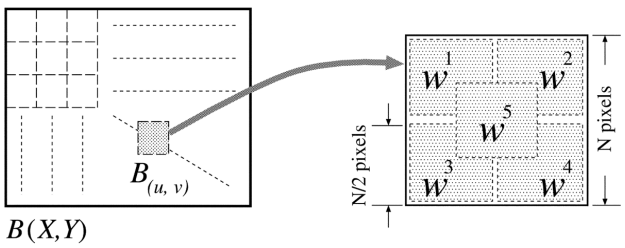


Fig. 3. Small windows in an image block.

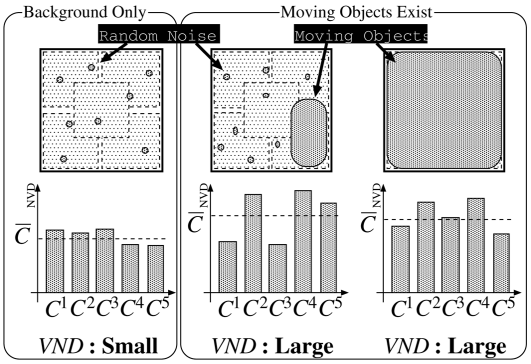


Fig. 4. Spatial configurations in a block and corresponding *VND* values.

(1) Background: In ideal noise-free conditions, the $C^j_{(u,v)}$ is 0 for each block, and $VND(\boldsymbol{i}_{(u,v)}) = 0$. When illumination is low and the effect of noise in the observed image is considerable, $C^j_{(u,v)}$ takes on a nonzero value. However, the noise is uniformly distributed, so the values are substantially equal to each other. Therefore, $VND(\boldsymbol{i}_{(u,v)})$, which is the noise dispersion, is a small value that does not depend on the illumination conditions.

(2) Combination of background and moving object: $C^j_{(u,v)}$ corresponding to the small windows in which a moving object is present has a large value, and other windows have a small value. Consequently, the dispersion $VND(\boldsymbol{i}_{(u,v)})$ has a large value.

(3) Moving object: As long as the moving object does not have the same texture as the background, the values of $C^j_{(u,v)}$ are randomly large. Consequently, $VND(\boldsymbol{i}_{(u,v)})$ increases.

As described above, the existence of a moving object can be determined by the magnitude of $VND(\boldsymbol{i}_{(u,v)})$.

2.4. Background subtraction based on the normalized vector distance

In the discussion up to this point, two invariable characteristic amounts $ND(\boldsymbol{i}_{(u,v)})$ and $VND(\boldsymbol{i}_{(u,v)})$ were obtained for varying illumination. These characteristic amounts are integrated and the following postulate is introduced in order to detect moving objects.

[Postulate 2] $ND(\boldsymbol{i}_{(u,v)})$ and $VND(\boldsymbol{i}_{(u,v)})$ follow the normalized distributions (mean: $m_{ND}\boldsymbol{i}^B_{(u,v)}$; dispersion $v_{ND}\boldsymbol{i}^B_{(u,v)}$) and (mean: $m_{VND}\boldsymbol{i}^B_{(u,v)}$; dispersion $v_{VND}\boldsymbol{i}^B_{(u,v)}$), respectively.

According to this postulate, the two-dimensional vector $(ND(\boldsymbol{i}^B_{(u,v)}), VND(\boldsymbol{i}^B_{(u,v)}))$ follows the two-dimensional normalized distribution [Fig. 5(a)]. The normalized

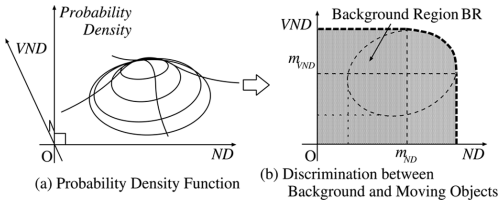


Fig. 5. Object detection based on NVD.

distribution $|\widetilde{\boldsymbol{i}^B_{(u,v)}}|$, that is, the background scene, is determined by the intensity of the illumination.

Figure 5(b) shows the discrimination border between the moving object and the background. If $(ND(\boldsymbol{i}^B_{(u,v)}), VND(\boldsymbol{i}^B_{(u,v)}))$ calculated for the block is within the shaded area BR, the area is the background, and areas not in the shaded area are determined to be areas that include a moving object. BR is defined below.

$$lk(ND(\boldsymbol{i}^B_{(u,v)}), VND(\boldsymbol{i}^B_{(u,v)})) > TH1 \quad (6)$$

$$\text{or} \quad ND(\boldsymbol{i}^B_{(u,v)}) < m_{ND}(\boldsymbol{i}^B_{(u,v)}) \quad (7)$$

$$\text{or} \quad VND(\boldsymbol{i}^B_{(u,v)}) < m_{VND}(\boldsymbol{i}^B_{(u,v)}) \quad (8)$$

In the formulas, the function *lk* is a likelihood based on the two-dimensional normalized distribution, and *TH*1 is the threshold value determined by $\sqrt{v_{ND}(\boldsymbol{i}^B_{(u,v)})}$.

2.5. Performance evaluation

A computational test was carried out in order to examine the effectiveness of the methods described to this point. The results are shown in Fig. 6.

The image used in the test was a grayscale image taken indoors with a fixed camera that was provided with a fluorescent light and was capable of controlling the combination of illumination level and lighting. The block size was set to 16×16.

In the test, the object detection results were compared by using the three methods described below.

- R_{ND}: Result of using a fixed threshold value with respect to $ND(\boldsymbol{i}_{(u,v)})$ with no consideration given to variance in the normalized vector distance due to noise.
- $R_{NDnoise}$: The varying illumination of the background scene was estimated by following a simple linear model* for varying illumination, and the threshold value with respect to $ND(\boldsymbol{i}_{(u,v)})$ was de-

*Varying illumination in $|\boldsymbol{i}_{(u,v)}| = \alpha_{(u,v)}|\boldsymbol{b}_{(u,v)}|$ is expressed by $\alpha_{(u,v)} = k_1u + k_2v + k_3$, and k_i (i = 1, 2, 3) is determined so that the difference in the brightness between the observed image and the estimated image is minimum.

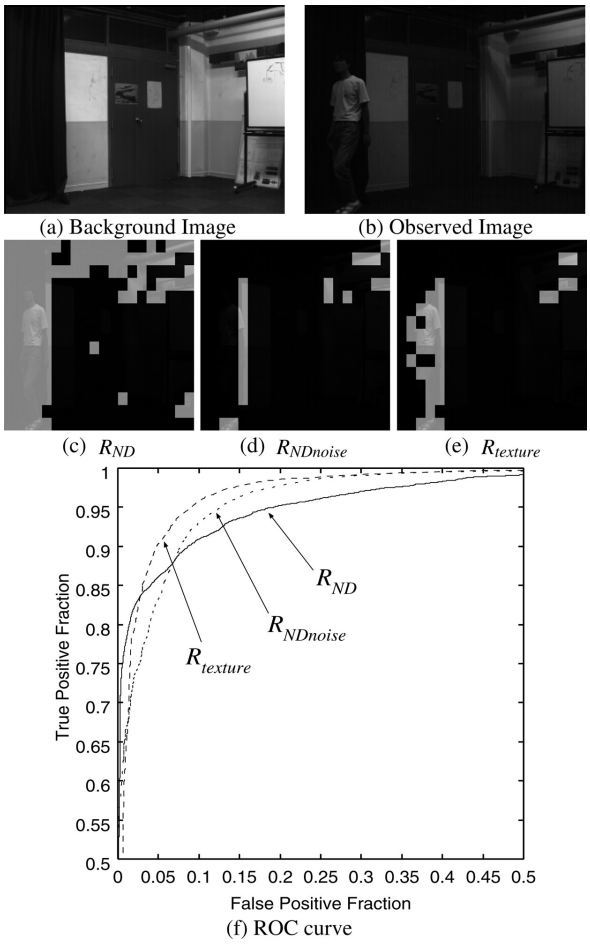


Fig. 6. Performance evaluation of the NVD-based method.

termined with consideration given to the variation of the normalized vector distance due to noise.

- $R_{texture}$: The method proposed in Section 2.4. This method uses the result of estimating $|\widetilde{\boldsymbol{i}^B_{(u,v)}}|$ by employing the same simple linear model described above.

In Fig. 6, (a) is the background image, (b) is the observed image under low illumination, and (c), (d), and (e) are the results R_{ND}, $R_{NDnoise}$, and $R_{texture}$, respectively, of processing (b). The difference in the detection results is conspicuous in the black ceiling, curtains, and other dark areas. In the R_{ND} method, misdetections are conspicuous due to the effect of noise; and in the $R_{NDnoise}$ method, detection omissions are conspicuous because the threshold value is set so that the detection sensitivity is lowered (Fig. 2). In contrast, in the $R_{texture}$ method, detection accuracy is improved because the evaluation includes spatial characteristics.

An ROC curve obtained in the manner described below is shown in Fig. 6(f) in order to quantitatively compare these three results. First, before the evaluation, the area of the moving object is given by hand precisely. Then, the result of averaging the detection ratio in each frame across the entire video image is recorded at a certain threshold value. Based on this result, an ROC curve is obtained by plotting the variation of the detection ratio while varying the threshold value. In the ROC curve, the vertical axis is the ratio at which the object is correctly detected (True Positive), and the horizontal axis is the ratio at which the background is mistakenly detected as an object (False Positive). The ROC curve in Fig. 6(f) shows the variation in the mean detection ratio under varying illumination, and it is clear that the detection accuracy of the proposed method is higher than with other methods.

A simplified $R_{NDnoise}$ has been implemented and evaluated by Toyama and colleagues [8] as well. They demonstrated robustness with respect to varying illumination, and the results shown in Fig. 6 confirm this robustness.

3. Background Subtraction Based on Estimation of Illumination Conditions

If the detection method described in Section 2 is used, robust object detection can be carried out in varying illumination, but there are drawbacks in that

- the intensity $|\widetilde{\boldsymbol{i}^B_{(u,v)}}|$ of the illumination of the background scene must be estimated in order to adaptively change the threshold value, and
- a moving object cannot be detected when both the background and the object have the same texture, and when both the background and object have a textureless, uniform brightness distribution.

The method described in this section solves these problems by detecting a moving object in the order of the following steps:

(1) estimating the illumination conditions of the observed image, and

(2) normalizing the brightness value with respect to varying illumination by using the estimation result and then carrying out background subtraction.

Here, the information required for detecting moving objects on the basis of the normalized vector distance described in the previous section can be obtained by using the estimation routine in step (1). Furthermore, in the method in which the normalized vector distance is used, the presence of an object and varying illumination cannot be

identified when the background and object have the same texture, but the method described in this section estimates the luminance of the background scene, making object detection possible in such a case.

On the other hand, object detection based on the normalized vector distance plays an important role, as described below, in estimating illumination conditions. Thus, the two methods described in this research paper work in a complementary fashion, and mutually increase accuracy.

3.1. Varying brightness model under varying illumination

Proposed in Ref. 9 is an Illumination Cone model that expresses variation in the brightness value due to varying illumination, and the following postulate is introduced. [Postulate 3]

- The surface of the object is a perfect diffusion surface.
- The objects are convex and shadows are not produced.
- All light sources are at an infinite point.

Let the vectors $\{\boldsymbol{i}_1, \boldsymbol{i}_2, \ldots, \boldsymbol{i}_n\}$ represent an image taken under different illumination levels. According to the Illumination Cone model, these vectors are distributed in a subspace in the form of up to a three-dimensional cone in an M^2-dimensional space. This subspace is defined by three eigenvectors $\boldsymbol{i}_{eigen1}$, $\boldsymbol{i}_{eigen2}$, and $\boldsymbol{i}_{eigen3}$. That is, the vector $\boldsymbol{i}_{any}$ corresponding to an image taken under any illumination conditions is given by

$$\boldsymbol{i}_{any} = a_1\boldsymbol{i}_{eigen1} + a_2\boldsymbol{i}_{eigen2} + a_3\boldsymbol{i}_{eigen3} \qquad (9)$$

In the expression, $a_k = \boldsymbol{i}_{any} \cdot \boldsymbol{i}_{eigenk}$ $(k = 1, 2, 3)$.

A test using an image taken indoors with a wide angle of view was carried out to confirm the validity of the varying illumination model. This is because the postulate described above holds up in narrow areas such as a human face, but the postulate is not necessarily satisfied in an image taken with a wide angle of view.

A fixed-viewpoint pan-tilt-zoom camera [1, 2] was used in order to obtain an indoor panoramic image. With this camera, images taken with different pan and tilt angles can be seamlessly combined to obtain a panoramic image. Figure 7 is an example of a panoramic image taken by varying the illumination intensity and the lighting pattern.

Objects in these scenes include a whiteboard with a reflecting surface, a mannequin and the shadow of a chair, and objects illuminated by a nearby light source, and the environment is one in which postulate 1 does not necessarily hold.

Fig. 7. Images taken under varying illumination.

The eigenvalues and eigenvectors were calculated from the observed image vectors $\{\boldsymbol{i}_1, \boldsymbol{i}_2, \ldots, \boldsymbol{i}_n\}$ by using principal component analysis. Such analysis entails first subtracting the average vector $\boldsymbol{i}_{avr}$ from the vectors, and calculating the eigenvalues and eigenvectors of the covariance matrix of $\{\boldsymbol{i}_1 - \boldsymbol{i}_{avr}, \boldsymbol{i}_2 - \boldsymbol{i}_{avr}, \ldots, \boldsymbol{i}_n - \boldsymbol{i}_{avr}\}$. This is because in an actual image, the principal components of an image vector distribution are often located at a distance from the origin.

Figure 8 shows the eigenvalues calculated from the images in Fig. 7. In a situation in which the postulate does not necessarily hold, the three eigenvalues have taken on considerably large values.

Figure 9 shows the eigenimages corresponding to the eigenvalues, respectively.

These test results show that the Illumination Cone model is effective in a real world indoor scene. The effectiveness of this model is described in greater detail in Section 6.

3.2. Method for detecting moving objects

Based on these tests, the illumination conditions of the observed image are estimated and background subtraction is carried out without the effect of varying illumination.

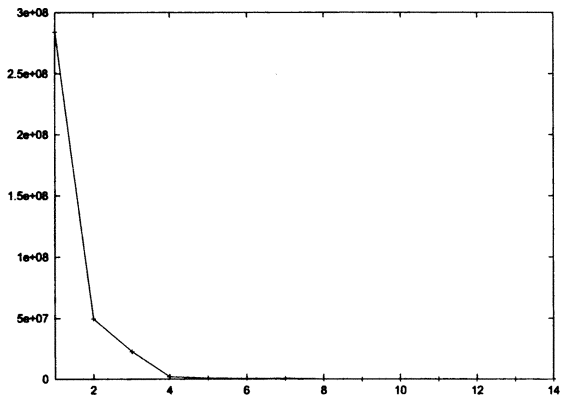


Fig. 8. Eigenvalues (vertical axis: magnitude; horizontal axis: index of eigenvalues).

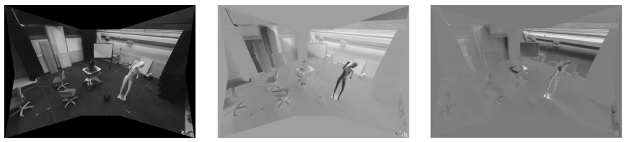


Fig. 9. Eigenimages.

(1) Background images are taken in various illumination conditions in order to configure the background scene.

(2) Principal component analysis is performed on the background images to obtain eigenvectors $\boldsymbol{i}_{eigen1}$, $\boldsymbol{i}_{eigen2}$, and $\boldsymbol{i}_{eigen3}$.

(3) The coefficient vector $\boldsymbol{a} = (a_1\ a_2\ a_3)^t$ for obtaining the background image of the observed point in time is calculated for the observed image $\boldsymbol{i}$ (image being processed) by using a generalized inverse matrix:

$$\boldsymbol{a} = (E^t E)^{-1} E^t (\boldsymbol{i} - \boldsymbol{i}_{avr}) \tag{10}$$

In the expression, the matrix E is defined as $E = |\boldsymbol{i}_{eigen1}\ \boldsymbol{i}_{eigen2}\ \boldsymbol{i}_{eigen3}|$ by using three eigenvectors.

(4) The image vectors in the estimated illumination conditions are given as follows:

$$\widetilde{\boldsymbol{i}} = a_1 \boldsymbol{i}_{eigen1} + a_2 \boldsymbol{i}_{eigen2} + a_3 \boldsymbol{i}_{eigen3} + \boldsymbol{i}_{avr} \tag{11}$$

(5) Subtraction for each pixel is carried out between $\widetilde{\boldsymbol{i}}$ and $\boldsymbol{i}$ to detect a moving object.

This algorithm is based on the premise that the area occupied by the moving object in the image is sufficiently small. Reference 10, in which the same method is used, demonstrates how small objects can be detected outdoors. However, a moving object often occupies a considerable area in an indoor image, and in such a case, estimation of the illumination conditions of the observed image, that is, the derivation of a_k ($k = 1, 2, 3$), is markedly affected by the moving object. Figure 10 demonstrates this fact.

Fig. 10. Stability of the illumination estimation. (a) An observed image that includes a moving object (human); (b) the estimation residual obtained using the entire image plane of the observed image $(a_1\ a_2\ a_3)$ = (9303.7 1013.0 –1363.3), mean residual: 2.49; and (c) the residual obtained when estimating without inclusion of the moving object $(a_1\ a_2\ a_3)$ = (7961.1 540.4 –434.9), mean residual: 1.23.

In the diagram, (a) shows an observed image that includes a moving object (human), (b) shows the residual of the result of estimating the illumination conditions using the entire observed image, and (c) shows the residual of the result of estimating the illumination conditions from the observed image without including the moving object ("the residual" in each case is increased fourfold). The values given in the caption are the mean residuals of the estimated coefficient a_k (k = 1, 2, 3), the observed image in the background area, and the estimated image. In (b), the mean residual is not considerable, but pixel values with a difference of several tens between 0 and 255 occur locally, causing misdetections. Thus, the moving object must be removed in order to estimate the illumination conditions with good precision. In order to achieve this, proposed method is devised as a detection method based on the normalized vector distance as described above.

4. Background Subtraction with the Integration of the Two Methods

Two detection methods were described above, but both methods have drawbacks.

[Detection by normalized vector distance] The brightness $|\widetilde{\boldsymbol{i}}^B_{(u,v)}|$ of the image block of the background scene must be known. Even if the observed image block $I_{(u,v)}$ is occupied by the moving object, the illumination intensity within the block may vary from the value obtained in advance, so illumination variation in (u, v) at the observed point in time must be estimated. [Detection by estimating illumination conditions] In order to estimate the illumination conditions with greater precision, the area corresponding to the moving object must first be removed from the observed image.

The two detection methods are recursively carried out, as described below, in order to solve these drawbacks (Fig. 11).

Step (1): An image background is photographed under various illumination conditions to obtain eigenimages $\boldsymbol{i}_{eigen1}$, $\boldsymbol{i}_{eigen2}$, and $\boldsymbol{i}_{eigen3}$. Also, the median of the pixels is calculated from the background image under high illumination to obtain a median image $\boldsymbol{i}_{median}$.

Step (2): Let the observed image of the object to be processed be $\boldsymbol{i}$. First, the object is detected based on the normalized vector distance, with $\boldsymbol{i}_{median}$ as the background image. Here, since the illumination conditions of $\boldsymbol{i}$ are unknown, $|\widetilde{\boldsymbol{i}}^B_{(u,v)}| = |\boldsymbol{i}_{(u,v)}|$. In this expression, $|\boldsymbol{i}_{(u,v)}|$ is the magnitude of the observed image vector in the block (u, v). Since the image $|\widetilde{\boldsymbol{i}}^B_{(u,v)}|$ is used solely to determine the threshold value, the object can be roughly detected even with this type of approximation.

Step (3): The pixels contained in the block with the moving object are removed from $\boldsymbol{i}$, $\boldsymbol{i}_{eigen1}$, $\boldsymbol{i}_{eigen2}$, and $\boldsymbol{i}_{eigen3}$.

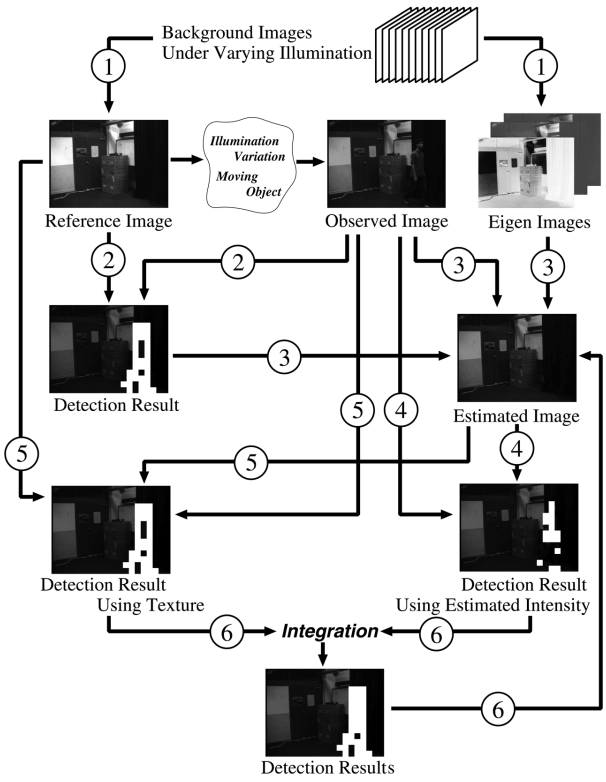


Fig. 11. Overview of the proposed method.

The illumination conditions are then estimated with respect to the remaining partial vectors to obtain the coefficients a_1, a_2, and a_3. Using these coefficients, the background image $\tilde{\boldsymbol{i}}$ that has been normalized to the illumination conditions of the observed point in time is calculated via Eq. (11).

Step (4): The difference in brightness between the observed image and the estimated background image is calculated and the moving object is detected in block units. Here, assume a moving object is present when the block (u, v) satisfies the condition $|\widetilde{\boldsymbol{i}_{(u,v)}}| - |\boldsymbol{i}_{(u,v)}|| > TH2$. Here, $TH2$ is the threshold value determined on the basis of the error in the estimation routine.

Step (5): Using the estimation results, object detection based on the normalized vector distance can be carried out again by letting $|\boldsymbol{i}^B_{(u,v)}| = |\widetilde{\boldsymbol{i}_{(u,v)}}|$.

Step (6): An OR operation is performed for each block with respect to the two detection results obtained in steps (4) and (5) to arrive at the final result. The estimation routine of step (3) is carried out again using this result. The processing up to this point is repeated until the estimation result is converged and the detection result stabilizes.

5. Experimentation

Figure 12 shows the median image $\boldsymbol{i}_{median}$, which is used as the background image.

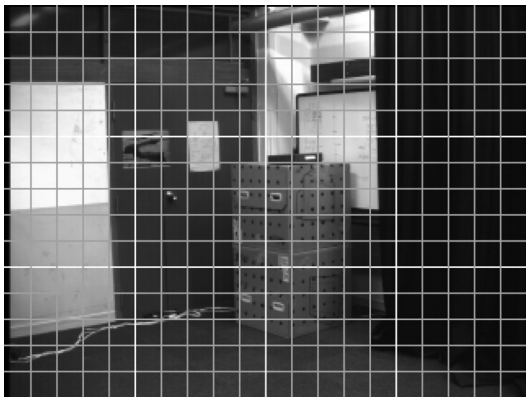


Fig. 12. Reference background image.

In the experiment, a grayscale image with a resolution of 256 and a size of 320 × 240 was used. In the experiment environment, the quantity of light could be continuously varied, and ceiling illumination was provided in which the lighting pattern could be selected. The experiment was carried out using an image taken in the above-described environment, with a person as the moving object.

The three methods that were compared for evaluating the performance are as follows.

$R_{texture}$: The detection results are based on the normalized vector distance described in step (2) of Section 4.

$R_{intensity}$: The detection results are based on the estimation of illumination conditions described in Section 3.2.

$R_{integrate}$: The detection results are obtained by integrating the two methods described in Section 4. Here, steps (3) to (6) are carried out only once.*

Note that the eigenimages needed to obtain $R_{intensity}$ and $R_{integrate}$ are calculated from 13 background images under different illumination conditions. The processing speed for obtaining $R_{integrate}$ is about three images per second.†

Figure 13 shows the ROC curve obtained using the image under high illumination.

Figure 14 shows the detection results in a frame. In the diagram, (a) is the correct area of the moving object, (b) is $R_{texture}$, (c) is $R_{intensity}$, and (d) is $R_{integrate}$. The results of each are obtained from the "optimum" threshold value. As used herein, the term "optimum" refers to the threshold value found by weighting the distance from (True Positive, False Positive) = (1, 0) on the ROC curve and taking the minimum value. The weighting was carried out with a True

*Because two threshold values $TH1$ and $TH2$ exist in this method, first, the point at which the detection ratio is most accurate is found by varying $TH1$ while $TH2$ is fixed. Next, the point at which the detection ratio is most accurate is found again by varying $TH2$. An ROC curve can be found by connecting such points.

†PC with a Pentium II at 400 MHz × 2.

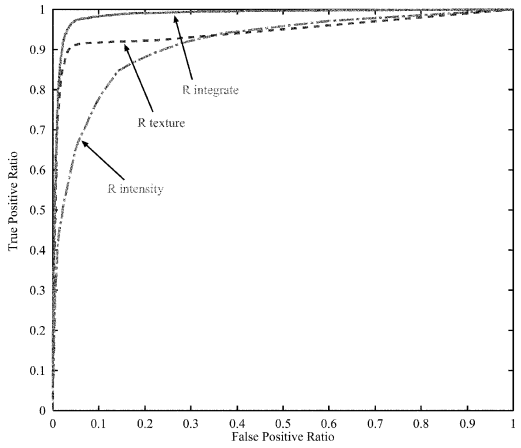
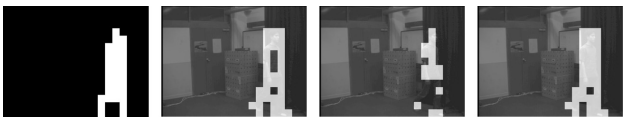


Fig. 13. Performance under high illumination.



(a) (b) (c) (d)

Fig. 14. Detected foreground objects under high illumination.

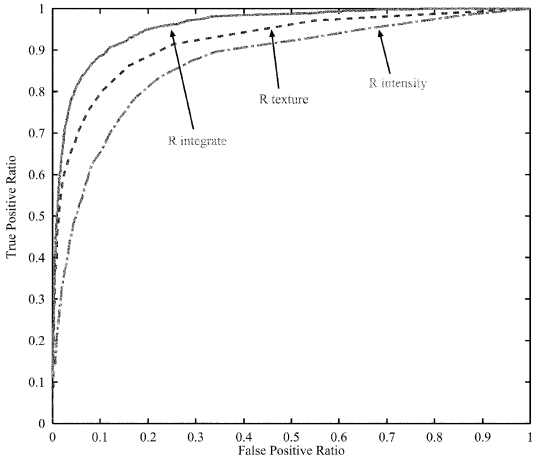


Fig. 15. Performance under low illumination.

(a) (b) (c) (d)

Fig. 16. Detected foreground objects under low illumination.

Positive : False Positive ratio of 3:1.* Figures 15 and 16 similarly show the results of experimentation under low illumination.

It is clear from both of these results that $R_{integrate}$, which is the result obtained via the proposed method, is considerably better than $R_{texture}$ and $R_{intensity}$, which are the results of the other methods. The reason for the poor detection ratio of $R_{intensity}$ is that the moving object occupies a large area and the accuracy of estimation is reduced. It is clear that the former is better when the results of high illumination are compared with low illumination. This is due to the fact that the SNR of the observed image is high when illumination is high.

6. Conclusion

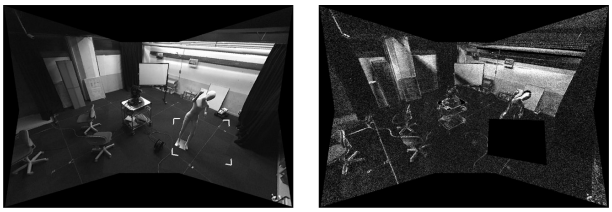
This research paper proposes a method for detecting moving objects using robust background subtraction under varying illumination. Two detection methods were described first. One method is based on the normalized vector distance defined for image blocks that correspond to the input image and the background image, and the other method is based on the estimation of the illumination conditions by using eigenimage analysis. Robust detection of moving objects under varying illumination was brought about by integrating these two methods, and the effectiveness of the proposed method was empirically demonstrated.

The method of estimating the illumination conditions must be changed when expanding the method proposed in this paper to detect a moving object in wider scenes by using an active camera. This is because various types of locally varying illumination are observed due to the illumination conditions and geometric configurations of objects in the real world. To confirm this, the following two experiments were carried out.

Figure 17(a) shows a panoramic image under certain illumination conditions. Square-shaped areas indicated by the four corners with broken lines were removed to estimate the varying illumination and obtain the coefficients a_k (k = 1, 2, 3). Next, the entire image under illumination conditions estimated using Eq. (11) was generated. The difference between this image and the observed image (a) is shown in (b). The differences in the square-shaped areas are few, and the local varying illumination can be correctly estimated, but in other areas the errors are greater in magnitude.

Next, the results of eigenimage analysis in block units are shown. Here, the rectangular area at the center of the panoramic image is divided into 15×7 blocks to find the eigenvectors. Figure 18 shows the dimensionality, that is,

*The goal was that the ratio in which the background is erroneously detected as the object be reduced to one-third the object detection leakage.



Observed image Error image

Fig. 17. Variations of local illumination conditions.

the number of dominant eigenvalues for a block. The geometric characteristics of the object surface whose dimensions are within the blocks are clearly observed. More specifically, one-dimensional areas correspond to flat surfaces (floors and walls), two-dimensional areas correspond to cylindrical shapes (chair legs and wall edges), and three-dimensional areas correspond to curved surfaces (mannequins and other complex surfaces). These results show that the type of object surface in a three-dimensional space can be classified using eigenimage analysis for each block, and the locally varying brightness value can be modeled on the basis of varying illumination.

The authors are currently researching systems for dynamically performing background subtraction using fixed-viewpoint pan-tilt-zoom cameras. In this area of research, locally varying illumination conditions are modeled by using eigenimage analysis for each block, and the three-dimensional characteristics of objects can be simultaneously acquired.

The authors expect that the precision of the proposed method can be improved by using color information. Even when objects that vary dynamically are present in the background, the variations of the object can be analyzed using spatial and temporal constraint conditions, as in Ref. 8. Proposed in Ref. 11, for example, is a method for modeling the variations of background objects on the basis of temporal correlations.

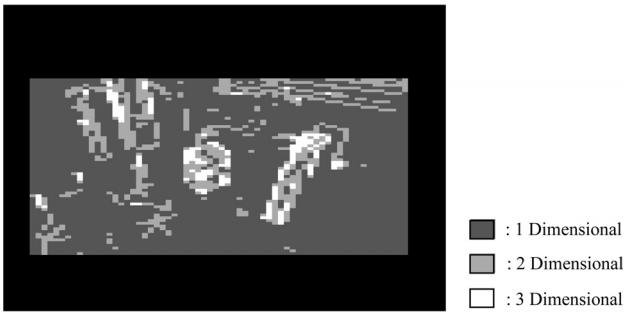


Fig. 18. Number of dominant eigenvalues for a block.

REFERENCES

1. Wada T, Matsuyama T. Appearance sphere: Background model for pan-tilt-zoom camera. Proc ICPR, Vol. A, p 718–722, 1996.
2. Matsuyama T. Cooperative distributed vision—Dynamic integration of visual perception, action, and communication. Proc Image Understanding Workshop, p 365–384, 1998.
3. Nakai H. Robust object detection using a-posteriori probability. IPSJ SIG-CV, Vol. 1994, No. 081, p 1–8, 1994. (in Japanese)
4. Grimson WEL, Stauffer C, Romano R, Lee L. Using adaptive tracking to classify and monitor activities in site. Proc CVPR, p 22–29, 1998.
5. Nagaya S, Miyatake T, Fujita T, Ito W, Ueda H. Moving object detection by time-correlation-based background judgment method. IEICE Trans Inf Syst Pt 2 1996;J79-D-II:568–576. (in Japanese)
6. Kagehiro T, Ohta Y. Acquisition and adaptive update of automatic background images from motion image sequence. Proc Meeting on Image Recognition and Understanding (MIRU) '94, Vol. II, p 263–270. (in Japanese)
7. Takatoo M, Kitamura T, Kobayashi Y. Vehicles extraction using spatial differentiation and subtraction. IEICE Trans Inf Syst Pt 2 1997;J80-D-II:2976–2985. (in Japanese)
8. Toyama K, Kramm J, Brumitt B, Meyers B. Wallflower: Principles and practice of background maintenance. Proc ICCV, p 255–261, 1999.
9. Belhumeur PN, Kriegman DJ. What is the set of images of an object under all possible lighting conditions? Proc CVPR, p 270–277, 1996.
10. Oliver N, Rosario B, Pentland A. A Bayesian computer vision system for modeling human interactions. Proc Int Conf on Vision Systems, p 255–272, Gran Canaria, Spain, 1999.
11. Matsuyama T, Ohya T, Habe H. Background subtraction for non-stationary scenes. Proc ACCV, p 662–667, 1999.

APPENDIX

Derivation of Theorem 1

The observed image block (size: $N \times N$) for only the background scene with added noise n_i is postulated in the following expression:

$$\left\{ \boldsymbol{i}^{B}_{(u,v)} \right\}_i = \widetilde{\boldsymbol{i}^{B}_{(u,v)}} + \boldsymbol{n}_i \qquad \text{(A.1)}$$

Hereafter, the variable i for expressing the observed data number is omitted. Here, it is assumed that the elements n_k of $\boldsymbol{n}$ each independently follow the normalized distribution of the mean 0 and the standard deviation σ, so the probability density function $f_n(\boldsymbol{n})$ that n follows is given as

$$f_n(\boldsymbol{n}) = \frac{1}{(2\pi\sigma^2)^{\frac{N^2}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{N^2} n_k^2 \right) \qquad \text{(A.2)}$$

In this case, the approximation of $ND_{(u,v)}$ shown in Fig. A.1 is used.

More specifically, consider plane P, which is perpendicular to $\widetilde{\boldsymbol{i}^{B}_{(u,v)}}$, and orthogonally project the noise vector $\boldsymbol{n}$ to P to obtain the vector $\boldsymbol{n}'$. Consider plane Q, which is parallel to P and is separated by distance 1 from the origin. The distance ND' between the intersection with the vector $\widetilde{\boldsymbol{i}^{B}_{(u,v)}} + \boldsymbol{n}'$ and the intersection with the vector $\widetilde{\boldsymbol{i}^{B}_{(u,v)}}$ is given by

$$ND' = \frac{1}{|\widetilde{\boldsymbol{i}^{B}_{(u,v)}}|} |\boldsymbol{n}'| \qquad \text{(A.3)}$$

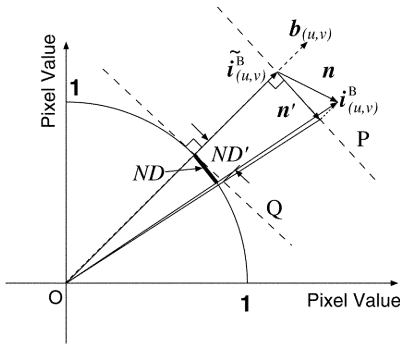


Fig. A.1. Approximation for calculating the effect of noise with relation to the normalized vector distance $ND_{(u,v)}$.

This is taken as the approximation of the normalized vector distance ND, and the probability density function that ND' follows is given in the procedure described below.

(1) Derive the probability density function that $|\boldsymbol{n}'|$ will follow.

(2) Substitute the result for ND' which approximates the normalized vector distance.

(1) Because the noise vector is isotropic, let the vector in which $\boldsymbol{n}$ is projected to the plane of $n_{N^2} = 0$ be $\boldsymbol{n}''$, and obtain the probability density function that $\boldsymbol{n}''$ will follow by using the expression

$$f_{n''}(\boldsymbol{n}'') = \frac{1}{(2\pi\sigma^2)^{\frac{N^2-1}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{N^2-1} n_k^2 \right)$$

Using the fact that the surface area of the $N^2 - 1$ dimensional spherical surface with the radius r in the N^2 dimensional space is $[2\pi^{(N^2-1)/2}/\Gamma((N^2-1)/2)]r^{N^2-2}$, and $|\boldsymbol{n}''|^2 = \Sigma_{k=1}^{N^2-1} n_k^2$, let $|\boldsymbol{n}''| = r$, and the probability density function that $r(r > 0)$ will follow will be given by

$$f_r(r) = \frac{2r^{N^2-2}}{(2\sigma^2)^{\frac{N^2-1}{2}} \Gamma(\frac{N^2-1}{2})} \exp\left(-\frac{r^2}{2\sigma^2} \right)$$

Consequently, the probability density function that $|\boldsymbol{n}'|$ will follow is

$$f_{n'}(|\boldsymbol{n}'|) = \frac{2|\boldsymbol{n}'|^{N^2-2}}{(2\sigma^2)^{\frac{N^2-1}{2}} \Gamma(\frac{N^2-1}{2})} \exp\left(-\frac{|\boldsymbol{n}'|^2}{2\sigma^2} \right)$$

Next, by transforming the variables of Eq. (A.3) for the calculated probability density function $f_{n'}(|\boldsymbol{n}'|)$, the probability density function that $d = ND'$ $(d > 0)$ will follow can be obtained in the form of the expression

$$f_{ND}(d) = \frac{2|\widetilde{\boldsymbol{i}^{B}_{(u,v)}}|^{N^2-1} d^{N^2-2}}{(2\sigma^2)^{\frac{N^2-1}{2}} \Gamma(\frac{N^2-1}{2})} \exp\left(-\frac{|\widetilde{\boldsymbol{i}^{B}_{(u,v)}}|^2 d^2}{2\sigma^2} \right)$$

Furthermore, the condition $|\widetilde{\boldsymbol{i}^{B}_{(u,v)}}| \simeq |\boldsymbol{i}^{B}_{(u,v)}|$ can be postulated to hold in the range of $|\boldsymbol{i}^{B}_{(u,v)}| >> |\boldsymbol{n}|$. Theorem 1 is obtained when the mean value and dispersion are calculated from this distribution.

AUTHORS (from left to right)




Takashi Matsuyama (regular member) received his B.E., M.E., and D.Eng. degrees in electrical engineering from Kyoto University in 1974, 1976, and 1980. He is currently a professor in the Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University. His research interests include knowledge-based image understanding, computer vision, 3D video, and dynamic human–machine interaction. He wrote about 100 papers and books including two research monographs, *A Structural Analysis of Complex Aerial Photographs* (Plenum, 1980) and *SIGMA: A Knowledge-Based Aerial Image Understanding System* (Plenum, 1990). He won six best paper awards from Japanese and international academic societies including the Marr Prize at ICCV'95. He is on the editorial boards of *Computer Vision* and *Image Understanding and Pattern Recognition*. He is now leading the five-year research project on Development of High Fidelity Digitization Software for Large-Scale and Intangible Cultural Assets, which is supported by the Ministry of Education, Culture, Sports, Science and Technology. He is a Fellow of the International Association for Pattern Recognition and a member of IEICE, the Information Processing Society of Japan, the Japanese Society for Artificial Intelligence, and the IEEE Computer Society.

Toshikazu Wada (regular member) received his B.E. degree in electrical engineering from Okayama University, M.E. degree in computer science from Tokyo Institute of Technology, and D.Eng. degree in applied electronics from Tokyo Institute of Technology in 1984, 1987, and 1990. He is currently a professor in the Department of Computer and Communication Sciences, Faculty of Systems Engineering, Wakayama University. His research interests include pattern recognition, computer vision, image understanding, and artificial intelligence. He received the Marr Prize at the International Conference on Computer Vision in 1995, the Yamashita Memorial Research Award from the Information Processing Society Japan (IPSJ), and the Excellent Paper Award from IEICE. He is a member of IEICE, IPSJ, the Japanese Society for Artificial Intelligence, and IEEE.

Hitoshi Habe (regular member) received his B.E. and M.E. degrees in electrical engineering from Kyoto University in 1997 and 1999. From 1999 to 2002, he was with Mitsubishi Electric Corporation. He then joined Kyoto University, and has been an assistant professor there since 2002. His research interests include computer vision and 3D image media processing. He is a member of the Information Processing Society of Japan and the IEEE Computer Society.

Kazuya Tanahashi received his B.E. degree in electrical engineering and M.Info. degree in intelligence science and technology from Kyoto University in 1998 and 2000 and joined NTT Data Corporation.